

BOIES SCHILLER FLEXNER LLP

David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsflp.com

Maxwell V. Pritt (SBN 253155)
Joshua I. Schiller (SBN 330653)
Joshua M. Stein (SBN 298856)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsflp.com
jischiller@bsflp.com
jstein@bsflp.com

Jesse Panuccio (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsflp.com

David L. Simons (*pro hac vice*)
55 Hudson Yards, 20th Floor
New York, NY 10001
(212) 909-7610
dsimons@bsflp.com

JOSEPH SAVERI LAW FIRM, LLP

Joseph R. Saveri (SBN 130064)
Cadio Zirpoli (SBN 179108)
Christopher K.L. Young (SBN 318371)
Holden Benon (SBN 325847)
Aaron Cera (SBN 351163)
Margaux Poueymirou (SBN 356000)
601 California Street, Suite 1505
San Francisco, California 94108
(415) 500-6800
jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
hbenon@saverilawfirm.com
acera@saverilawfirm.com
mpoueymirou@saverilawfirm.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
(323) 968-2632
mb@buttericklaw.com

CAFFERTY CLOBES MERIWETHER & SPRENGEL LLP

Bryan L. Clobes (*pro hac vice*)
135 S. LaSalle Street, Suite 3210
Chicago, IL 60603
(312) 782-4880
bclobes@caffertyclobes.com
asweatman@caffertyclobes.com

*Counsel for Individual and Representative Plaintiffs
and the Proposed Class. (additional counsel included below)*

**UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION**

RICHARD KADREY, et al.,

Individual and Representative Plaintiffs,

v.

META PLATFORMS, INC.,

Defendant.

Case No. 3:23-cv-03417-VC

**PLAINTIFFS' REPLY TO DEFENDANT
META PLATFORMS, INC.'S OPPOSITION
TO MOTION TO AMEND CASE
SCHEDULE**

Plaintiffs respectfully submit this reply to Defendant Meta Platform, Inc.’s (“Meta’s”) opposition to Plaintiffs’ renewed motion to amend the initial case management order. ECF No. 201.

Meta’s opposition fails to acknowledge the Court’s recognition that this case involves an “important societal issue” and thus summary judgment, including on the issue of fair use, should be decided on a “proper record.” ECF No. 184 (“Mot.”) at 1. The opposition also does not dispute that the existing and additional discovery Plaintiffs identified in their motion are important to the fair use issue, among others. Instead, Meta appears to believe that the record on which the Court decides such “important questions,” *id.*, should be one dictated by Meta alone, and that the record is sufficient so long as Meta produced *some* discovery on issues relevant to those questions, rather than *complete* discovery on *all* relevant issues. But that is not how litigation works, let alone in a case involving the important and novel societal issue of LLMs’ unlicensed use of copyrighted works on an unprecedented scale.

Tellingly, Meta also continues to play coy about the scope of summary judgment, despite the Court’s request that Meta “identify the issues that we need to . . . deal with at summary judgment.” Mot. at 1. Meta first draws a meaningless distinction between its refusal to stipulate that the five issues the Court assumed would be undisputed on summary judgment, *id.* at 2 n.2, are in fact undisputed, and the question of whether Meta will move for summary judgment on them. *See* ECF No. 201 (“Opp.”) at 2:2-16. And even that empty distinction falls away because Meta goes on to say “it may make legal arguments regarding whether ‘copying’ has occurred,” *id.*, making clear summary judgment—and the discovery needed to develop a proper record for the Court—plainly is *not* limited to fair use but “may,” according to Meta, include all liability issues. Meta also still refuses to confirm whether it will raise at summary judgment any affirmative defenses other than certain standing issues.

*Id.*¹

Finally, Meta does not address, let alone raise concerns with, Plaintiffs’ proposed amended case management schedule based on Meta’s confirmation that it “may” raise liability issues at

¹ If Meta intends to challenge any of the named plaintiffs’ standing, whether as to copyright ownership or otherwise, it should do so now—not at summary judgment aimed at liability and fair use.

summary judgment and will raise affirmative defenses besides fair use. *See* Mot. at 3.² For these reasons alone, Plaintiffs respectfully ask that the Court grant the Motion and amend the case management deadlines as set forth on page three of the Motion, with the close of fact discovery on January 24, 2025, and summary judgment and *Daubert* motions to be fully briefed by June 13, 2025.³ Plaintiffs also request the Court set a further case management conference for December 20, 2024.

* * *

As stated above, while Meta does not dispute that the existing and additional discovery Plaintiffs identify in the Motion are important to issues to be decided at summary judgment, Plaintiffs respond below to Meta’s statements on discovery and to correct some of Meta’s misstatements.

1. Discovery needed to make a proper fair-use record:

a. Llama Source Code and Technical Data: Meta states that Plaintiffs “incorrectly” assert that Meta has not produced source code and other technical data for *all* Llama models, while simultaneously admitting Meta has *not* produced data for *all* Llama models, including Llama 4 or the more recent Llamas 3.1 and 3.2, which Meta does not mention. As for Llama 3, Meta does not dispute that it has not produced any additional source code for that model, and says it only produced some of its training data. *See* Opp. at 2; Hartnett Decl., ¶¶ 4-5 (counsel’s understanding is that Meta made available source code for Llamas 1 and 2, “much of which was used for Llama 3,” and that it supplemented its production as to Llamas 1 and 2 with self-selected “pertinent” Llama 3 training data).⁴ Meta states elsewhere without explanation that unspecified datasets and technical data Plaintiffs identify are “irrelevant.” Opp. at 3-4. Meta also fails to mention it objected as recently as

² Plaintiffs decline to address Meta’s false *ad hominem* attacks, like “the case schedule now proposed by Plaintiffs has not flowed from a close assessment of the record and showing of need.” Opp. at 3.

³ Under this schedule, Plaintiffs believe the Court will remain the first in this District and others to address liability and fair use issues at summary judgment with respect to claims of infringement by LLMs.

⁴ Indeed, just this week around 2:45 p.m. on October 2, Meta “refreshed” the source code machines it made available to Plaintiffs, adding about 26,000 commits (changes to source code files), which almost doubled the number of commits it previously made available. Krein Decl., at ¶ 12. Meta also added for the first time this week over 9,000 pull requests with respect to each code repository, *id.* ¶ 13, which should have been produced in the normal course of discovery, like it has in similar cases, but it appears Meta did not search this non-custodial data in response to Plaintiffs’ discovery requests. And Meta still hasn’t provided code and technical data for the entire relevant period. *Id.*, ¶¶ 14-17.

last week to discovery about all models of Llama and all datasets of copyrighted works that Meta used to train its LLMs, *see* Pritt Reply Decl., Ex. A, ¶¶ 5 (objecting to discovery of all Llama models), 9 (objecting to discovery of pirated datasets other than Books3), despite documents showing Meta pirated works from multiple sources and did so repeatedly, *see* Benon Decl., Ex. A. Further, Meta says “the parties agreed that Plaintiffs did not want, and Meta did not need to produce, additional Llama 3 training data.” Opp. at 3; Hartnett Decl., ¶ 5. That is simply false. *See* Benon Decl., ¶¶ 3-6; Fishman Decl., ¶ 4. Discovery into these issues is relevant both to infringement and fair use factor one. *See* ECF No. 133, ¶ 83 (class definition includes all Llama models and all training datasets). Meta’s copyright mitigation techniques are also evidence of direct infringement and of Meta’s myriad uses of protected works in connection with its LLMs, and therefore relevant both to liability and fair use, which Meta fails to address other than citing to the Court’s motion to dismiss order. Opp. at 4.⁵

b. Licensing Discovery: Meta fails to refute, or even engage with, Plaintiffs’ showing that discovery into Meta’s efforts to license training data for its LLMs is essential to the Court’s analysis of fair use factor four. Unable to respond on the merits (the paucity of documents produced on this topic speaks for itself), Meta instead seeks to minimize the importance of this discovery by relying on several incomplete or misleading statements. *First*, Meta claims to have “responded to Plaintiffs’ discovery requests regarding licensing” (Opp. at 2), but conspicuously omits it produced *zero documents* in response, presumably because none of Meta’s ESI custodians were on the Business Development team responsible for investigating licensing deals. (*See* Pritt Reply Decl., Ex. B., RFP 78).⁶ *Second*, Meta’s argument about recent changes to the deposition schedule ignores that Plaintiffs rescheduled or changed certain depositions because (a) they needed to stay within the current limit

⁵ Meta likely used training datasets comprising unlicensed, copyrighted works to teach its LLMs what is “copyrighted” versus in the public domain, including pirated works contained within Books3, LibGen, and the other shadow libraries from which documents show Meta sourced data (i.e., Anna’s Archive, Internet Archive, and Z-Lib, a repository Meta described as “banned”). Benon Decl., Ex. A. Thus, Meta’s “techniques” for preventing its models from outputting copyrighted material are evidence both of direct infringement and of Meta’s myriad “uses” of actual copies of copyrighted works, which are relevant to all of Meta’s claimed “fair uses” of those works.

⁶ Similarly, Meta responded to an earlier RFP on the same topic by refusing to produce documents (*see* Pritt Reply Decl., Ex. C, RFP 64), and Meta improperly limited its interrogatory response on this topic to any “executed written agreements concerning licensing or acquisition of training datasets,” thereby siloing off ongoing discussions or inchoate deliberations. *See id.*, Ex. D, ROG 5.

of 10 depositions, which Plaintiffs sought to increase; and (b) Meta refused to produce custodial documents for all deponents. Judge Hixson said he will *sua sponte* revisit Plaintiffs' request to add deponents and custodians should the Court modify the current schedule and extend the close of discovery. (ECF Nos. 146, 196.) *Third*, Meta's self-serving claim that "Meta's custodians . . . are those individuals most likely to possess responsive documents" (Opp. at 2-3) only further supports Plaintiffs' argument that Meta improperly excluded from discovery documents bearing directly on the potential market for licensing Plaintiffs' works given the lack of responsive documents produced from those custodians.

c. Privilege Issues: Meta states "no motion has been filed with Judge Hixson," Opp. at 5, but fails to tell the Court that is because Meta agreed to provide an amended privilege log on October 7 after Plaintiffs raised deficiencies with it and Judge Hixson ordered the parties to brief the issue by October 3. Because Meta's amended log could moot any dispute over its adequacy, the parties agreed to raise the issue, if still necessary, with Judge Hixson after Meta serves, and Plaintiffs review, its amended log. Meta also fails to advise the Court that it re-produced many documents it redacted for privilege this past weekend after Plaintiffs raised concerns about those redactions, and that it continues to redact business communications, including those without attorneys present. These issues must be resolved.

d. Fair Use Interrogatory and Advice of Counsel: Meta says it provided "fulsome responses" on September 30. Not so. The responses are deficient, and Plaintiffs will ask Meta to amend them.⁷

In addition to the foregoing issues, it has become apparent during the last few days that Meta's custodial searches were deficient—for example, Plaintiff asked Meta weeks ago whether it searched certain custodial data sources including text and WhatsApp messages. Meta did not respond. Then, just yesterday morning, a few hours before the deposition of Ahmad Al-Dahle, the head of Meta's

⁷ Meta's responses not only improperly limit the scope of the requests (e.g., to only some Llama models and to only one of many pirated datasets) but fail to provide complete and non-evasive answers. For example, Meta states it "does not *presently* intend to assert the advice of counsel defense," Pritt Reply Decl., Ex. A at 13 (emphasis added), and fails to provide any information concerning critical statements it makes like "any particular book has no independent value as training data and/or is interchangeable with countless other books," and it would be "impractical, if not impossible, for companies developing LLMs to attempt to negotiate licenses," *id.* at 12-13.

Generative AI efforts and one of Meta’s self-selected 10 document custodians, Meta produced—for the first time—WhatsApp communications between Mr. Al-Dahle and Meta’s CEO, Mark Zuckerberg. Meta also produced other responsive communications during the few days leading up to Mr. Al-Dahle’s deposition. The failure to produce these documents earlier calls into question both Meta’s self-selected and undisclosed search terms and whether it searched relevant non-custodial and custodial sources.⁸ Meta also incorrectly imaged documents it produced, which Judge Hixson ordered Meta to fix on October 1, see ECF No. 196 at 3, and Meta has failed to produce many hyperlinked documents or the entirety of comments in other produced documents.

2. Additional discovery needed for a proper record to decide issues at summary judgment:

Meta similarly fails to dispute the importance of the three additional categories of discovery Plaintiffs identified as required to develop a proper record: (1) financial information about Llama models’ revenue and profit impacts, (2) Meta’s state of mind, and (3) additional technical discovery. Instead, Meta points to similar or related discovery. *See* Opp. at 5 (citing Hartnett Decl.). But there is no actual dispute that Meta has *not* produced the type of information that would allow Plaintiffs (to say nothing of a jury) to understand fully what kind of revenues and profits are driven by Meta’s infringement. Nor is there a real dispute that Meta is sitting on troves of relevant documents from top AI decision-makers, including its CEO (whose deposition Judge Hixson ordered on September 24 to proceed, although Meta has still not provided a date despite Plaintiffs’ request that same day), that speak directly to purpose (fair use factor one) and willfulness (a key driver of copyright damages); that leaves Plaintiffs finding more commentary from Meta executives in the press than in produced documents. *See, e.g.*, Mot. at 1 (quoting a recent interview given by Meta’s CEO).

* * *

On the current record, Plaintiffs’ request to extend the discovery deadline and pre-trial schedule by three months is both reasonable and appropriate in such an important and novel case.

⁸ Because the ESI order in this case did not require the parties to exchange this information at the outset of this litigation, like in other similar cases, absent an agreement (or court order) to do so now, the parties must engage in discovery on discovery and serve discovery requests seeking the information.

Dated: October 4, 2024

Respectfully submitted,

BOIES SCHILLER FLEXNER LLP

By: /s/ David Boies

JOSEPH SAVERI LAW FIRM, LLP

Joseph R. Saveri (SBN 130064)
Cadio Zirpoli (SBN 179108)
Christopher K.L. Young (SBN 318371)
Holden Benon (SBN 325847)
Aaron Cera (SBN 351163)
601 California Street, Suite 1505
San Francisco, California 94108
(415) 500-6800
jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
hbenon@saverilawfirm.com
acera@saverilawfirm.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
(323) 968-2632
mb@buttericklaw.com

CAFFERTY CLOBES MERIWETHER & SPRENDEL LLP

Bryan L. Clobes (*pro hac vice*)
Alexander Sweatman (*pro hac vice*)
Mohammed Rathur (*pro hac vice*)
135 S. LaSalle Street, Suite 3210
Chicago, IL 60603
(312)-782-4880
bclobes@caffertyclobes.com
asweatman@caffertyclobes.com
mrathur@caffertyclobes.com

DICELLO LEVITT LLP

Amy Keller (*pro hac vice*)
Nada Djordjevic (*pro hac vice*)
James Ulwick (*pro hac vice*)
10 North Dearborn Street, Sixth Floor
Chicago, Illinois 60602

David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsflp.com

Maxwell V. Pritt (SBN 253155)
Joshua I. Schiller (SBN 330653)
Joshua M. Stein (SBN 298856)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsflp.com
jischiller@bsflp.com
jstein@bsflp.com

Jesse Panuccio (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsflp.com

David L. Simons (*pro hac vice*)
55 Hudson Yards, 20th Floor
New York, NY 10001
(914) 749-8200
dsimons@bsflp.com

(312) 214-7900
akeller@dicellosevitt.com
ndjordjevic@dicellosevitt.com
julwick@dicellosevitt.com

DICELLO LEVITT LLP
David Straite (*pro hac vice*)
485 Lexington Avenue
Suite 1001
New York, New York 10017
(646) 933-1000
dstraite@dicellosevitt.com

Counsel for Individual and Representative Plaintiffs and the Proposed Class